C-Grain Summarize Program

User Guide for usda-c-grain-sum

May / 2024

Table of Contents

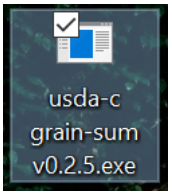
C-Grain

The C-Grain is used to rapidly take images from a large number of seeds or kernels and produce numeric output from those images. It has two main output formats: csv and xml. In the csv output files, information is listed for each kernel, with a number of measurements provided. In the xml output files, measurements are only listed per sample (composed of many kernels), and some machine learning results from each sample are also included.

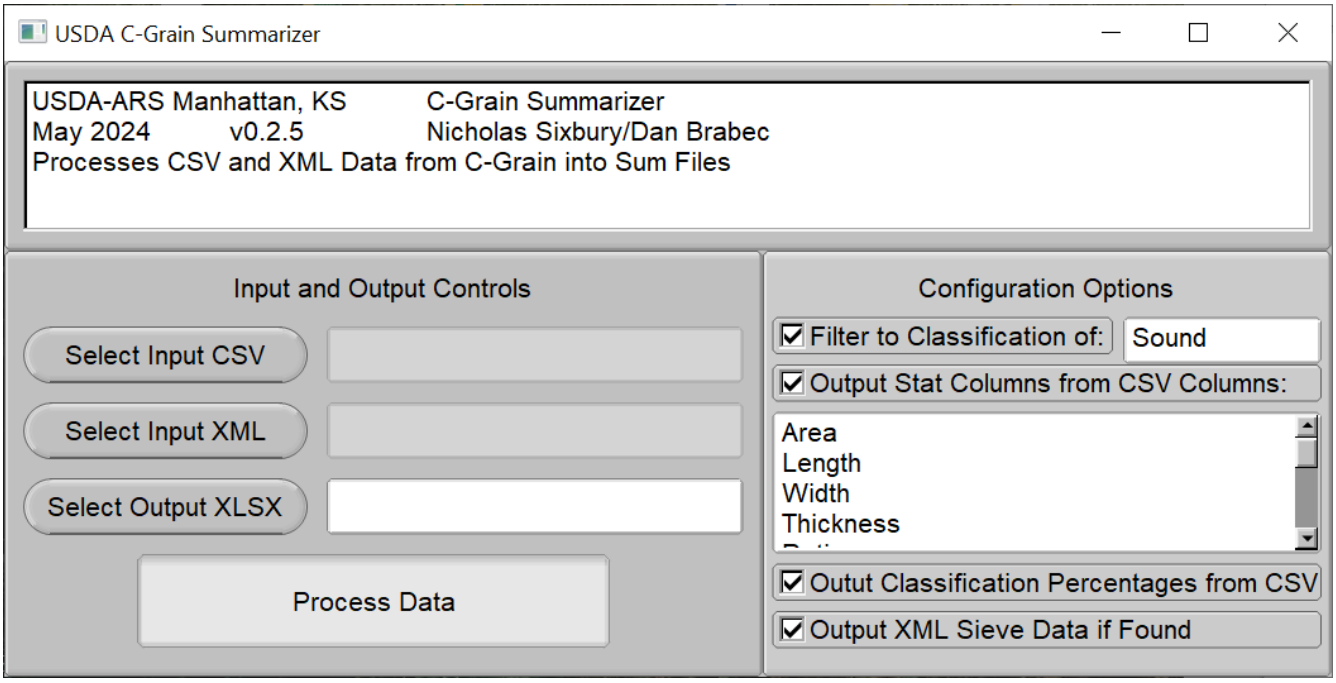
usda-c-grain-sum software

This software was written at USDA-ARS of Manhattan by Nicholas Sixbury in the spring and summer of 2024, developed for use by scientists Scott Bean and Rhett Kaufman. The purpose of the software is to extract the useful information from the csv and xml files created by the C-Grain in order to save time that would otherwise be spent manually processing the files in excel. The goal of the project is to create a summary of each sample, which contains a number of kernels. This summary is created as an excel file with each sheet holding different information on each sample.

How to use the summary software:



1. **Open the program**, by using the mouse to click on the icon.
2. Choose input files by clicking "**Select Input CSV**" and/or "**Select Input XML**"
3. Choose output file name by clicking "**Select Output XLSX**" or just typing a name in the box.
4. Check your preferred options.
5. Click "Process Data" to create the output file from the inputs you selected.



Advanced Options for Sum Software:

Reset the Configuration Settings

If, at any time, you wish to reset the configuration settings, simply right-click the label “Configuration Options” to the right. A dialog box will pop up asking if you’re sure you want to continue. Click yes, and another dialog box will appear, asking which preset you’d like. Select any option, and your configuration will be reset to the option you choose.

Change the Classification Filter for CSV Stat Columns

In order to enabled class filtering for csv stat columns, the checkbox labeled “Filter to Classification of” must be checked. By the default, the box to the right will say either “Sound” or “Sorghum”. The way this classification works is that it removes kernels with classes not specified from consideration. So, if “Sound” is specified, then csv stat columns will only consider kernels with a classification of “Sound”. If you wish to filter for multiple classes, such as “Sound”, or “Black-Germ”, then you can enter however many class names you wish, separated by the | character. If multiple classes are specified, then kernels with any of the classes specified will be considered. In order to determine what a kernel is classified as, the program looks for the column in the csv labelled “raw-filtered-as”. If this column cannot be found, then it will default to looking in the 6th column.

Change the Column Config for CSV Stat Columns

The csv stat columns output option calculates the average and standard deviation for several columns of data in the csv. The way this is configured is quite simple: If the text “Area” is placed in the column config, then the program will look for a column called “Area” in the csv. Each value in the config can be placed on a different line. If you wish to see a list of common columns names, simply click the config area, which lists all the column names, and press the F1 key.

Changing the Sample ID Grouping in the Output for CSV Files

By default, in order to determine the sample id of a particular kernel, the program looks for the column labelled “external-sample-id”, and if that column cannot be found, then the program defaults to using the 3rd column, which should have the label “external-sample-id” by default. If you wish to use a different column for grouping samples, simply ensure that the only column labelled “external-sample-id” is the one you wish to use. This can be easily done by editing the column header in the csv, though for long-term use, it is recommended that you contact the developer and request a new feature.

Troubleshooting:

Program displays a console window when opened:

The program currently prints some debugging information to the console while running. During normal operation, this debugging information can be safely ignored, and anything requiring your attention will be communicated through the graphical interface. You can safely minimize the console window to get it out of the way.

When opened, Program displays a dialog message saying something about config:

The program saves certain configuration information to a file in order to save your settings when the program is closed. The file is located in the same directory as the executable file for portability. If the config file cannot be found, the program assumes that you opening the program for the first time, so it will ask you if you want to apply a preset. If you don't want to bother with it, simply click "no", and a default config will be created. Once this process completes, and you close the program once, you shouldn't see the message again. If you continue seeing the message every time you open the program, this indicates that the application is encountered difficulties when saving the file. If this occurs, you can still use the program normally, but please contact the developer for support in order to resolve the issue.

After clicking "Process Output", Program displays series of Error dialogs about data:

The most likely cause of this problem is that one or both of the input files are not the correct type. When you click "Select Input CSV", a file dialog will open, which is filtered by default to only shown csv files. If you select a non-csv file anyways, the program will allow you to open it, but if the data within is not formatted like a csv, it won't work. The same issue can happen when selecting non-xml files for "Select Input XML". If you're sure that the file you selected is the right type, open the file in question and check if there are any discrepancies in the formatting.

After clicking "Process Output", Program displays series of Error dialogs and closes:

The most likely cause of this problem is an invalid output path name. Each operating system has certain characters that cannot be included in a filename. If you type your output filename into the box manually, then there is no check preventing you from creating an invalid output path, thus when the program tries to write to the path, it encounters an error. This can be fixed by not including illegal characters, which is made easier when you use the file dialog opened by "Select Output XLSX" to select your output file, as you will be warned when trying to create an invalid filename.